Efficient estimators for quantum instanton evaluation of the kinetic isotope effects: application to the intramolecular hydrogen transfer in pentadiene

Jiří Vaníček[1,2,*] and William H. Miller[1]

[1]*Department of Chemistry and Kenneth S. Pitzer Center for Theoretical Chemistry, University of California, Berkeley, CA 94720*

[2]*Simons Center for Systems Biology, Institute for Advanced Study, Princeton, New Jersey 08540*

(Dated: June 13, 2007)

Abstract

The quantum instanton approximation is used to compute kinetic isotope effects for intramolecular hydrogen transfer in *cis*-1,3-pentadiene. Due to the importance of skeleton motions, this system with 13 atoms is a simple prototype for hydrogen transfer in enzymatic reactions. The calculation is carried out using thermodynamic integration with respect to the mass of the isotopes and a path integral Monte Carlo evaluation of relevant thermodynamic quantities. Efficient "virial" estimators are derived for the logarithmic derivatives of the partition function and the delta-delta correlation functions. These estimators require significantly fewer Monte Carlo samples since their statistical error does not increase with the number of discrete time slices in the path integral. The calculation treats all 39 degrees of freedom quantum-mechanically and uses an empirical valence bond potential based on a modified general AMBER force field.

I. INTRODUCTION

The kinetic isotope effect (KIE) is defined as the ratio k_A/k_B of rate constants for two isotopologs A and B. Isotopologs A and B are two chemical species differing only by replacing an atom (or a group of atoms) in molecule A by its (their) isotope(s) in molecule B.

Kinetic isotope effects are among the main tools of chemical kinetics for determining mechanisms of complex chemical reactions or the extent of nuclear quantum mechanical effects in a simple reaction.[1] These effects are particularly strong for hydrogen transfer reactions with a high activation barrier or at low temperatures. However, unexpectedly large KIE's have recently been observed also in several enzymatic reactions at physiological temperatures.[2,3] These experiments imply that in some biochemical reactions, the hydrogen transfer proceeds via tunneling even at physiological temperatures, often with the help of thermally excited "promoting" vibrations that modulate the distance between the hydrogen donor and acceptor. These quantum effects have also been observed in several numerical simulations.[4–7] Most simulations include quantum effects by semiclassical (SC) approximations or by treating one or two degrees of freedom quantum-mechanically and the rest classically.

In a recent paper,[8] a different approach was developed, based on the quantum instanton approximation (QI) for the rate constant.[9] This approximation is similar to the semiclassical instanton theory,[10] but has an advantage that the Boltzmann operator is treated fully quantum mechanically rather than within the SC approximation. The quantum instanton theory thus incorporates all the tunneling, corner-cutting, and quantum-fluctuations effects correctly and is expected to overcome the quantitative deficiency of the SC instanton model. Indeed, several test applications have shown the QI theory to give accurate quantum rates over a wide temperature range, from the "deep" tunneling regimes at low temperatures to the regime of over-barrier dynamics at high temperature.[9,11–14] In Ref. 8 the QI was successfully applied to compute KIE's in a simple hydrogen exchange reaction. The calculation was based on a thermodynamic integration with respect to the mass of isotopes and the path integral representation of relevant quantities.

In this paper, this method is developed further: specifically, we derive very efficient estimators needed in polyatomic reactions. The advantage of these "virial" estimators is

that their error is independent of the number of discrete time slices in the path integral. We then test this improved method on the KIE in the [1,5]-sigmatropic hydrogen shift in *cis*-1,3-pentadiene.

The outline of the paper is as follows: The QI approximation for the KIE is described in Sec. II. The new and more efficient virial estimators are derived in Sec. III. In Sec. IV, the improved method is tested on the intramolecular hydrogen transfer in pentadiene. In particular, the virial estimators are compared to the thermodynamic estimators from Ref. 8. The importance of quantum effects are judged by increasing the number of imaginary time slices and by comparison with mixed quantum-classical calculations. Finally, Sec. V discusses the merits of the present approach and its possible extensions in future applications in real enzymatic reactions.

II. QUANTUM INSTANTON APPROXIMATION FOR THE KINETIC ISOTOPE EFFECTS

The quantum instanton (QI) approximation for the rate constant was introduced in Ref. 9. Here we summarize a simpler alternative derivation[15] described in Ref. 8. This derivation starts with the Miller-Schwartz-Tromp formula[16] for the thermal rate constant k,

$$k\,Q_r = \int_0^\infty dt\, C_{ff}\,(t)\,. \tag{2.1}$$

Here Q_r is the reactant partition function (per unit volume for bimolecular reactions) and $C_{ff}\,(t)$ is the symmetrized (and therefore real) flux-flux correlation function,

$$C_{ff}\,(t) = \mathrm{tr}\left(e^{-\beta\hat{H}/2}\hat{F}_a e^{-\beta\hat{H}/2} e^{i\hat{H}t/\hbar}\hat{F}_b e^{-i\hat{H}t/\hbar}\right) \tag{2.2}$$

with Hamiltonian operator $\hat{H}$ and quantum flux operators $\hat{F}_\gamma$ through dividing surfaces $\gamma = a, b$ (see Ref. 8). The QI expression for the rate follows by multiplying and dividing the integrand of Eq.(2.1) by the "delta-delta" correlation function $C_{dd}\,(t)$, assuming that $C_{ff}\,(t)\,/C_{dd}\,(t)$ varies slowly compared with $C_{dd}\,(t)$, and applying the steepest descent approximation to the resulting integral. Assuming further that the stationary-phase point is at $t = 0$, we obtain the QI thermal rate constant,

$$k_{\mathrm{QI}} = \frac{1}{Q_r} C_{ff}\,(0)\,\frac{\sqrt{\pi}}{2}\frac{\hbar}{\Delta H}. \tag{2.3}$$

Here ΔH is a specific type of energy variance,[11]

$$\Delta H = \hbar \left[\frac{-\ddot{C}_{dd}(0)}{2C_{dd}(0)} \right]^{1/2}, \tag{2.4}$$

where the delta-delta correlation function $C_{dd}(t)$ is defined[8,12] as

$$C_{dd}(t) = \mathrm{tr}\left(e^{-\beta\hat{H}/2} \hat{\Delta}_a e^{-\beta\hat{H}/2} e^{i\hat{H}t/\hbar} \hat{\Delta}_b e^{-i\hat{H}t/\hbar} \right), \tag{2.5}$$

with $\hat{\Delta}_\gamma$ being a generalized delta operator (see Ref. 8).

In applying the QI approximation to the KIE, it is useful to consider a continuous change of the isotope mass. If the two isotopologs are A and B, a real parameter $\lambda \in [0,1]$ can be defined such that

$$m_i(\lambda) = m_{A,i}(1-\lambda) + m_{B,i}\lambda \tag{2.6}$$

where $m_{A,i}$ and $m_{B,i}$ are the masses of the ith atom in the isotopologs A and B, respectively. Within the quantum instanton approximation (2.3), the KIE can be expressed as

$$\frac{k_{\mathrm{QI}}(0)}{k_{\mathrm{QI}}(1)} = \frac{Q_r(1)}{Q_r(0)} \times \frac{\Delta H(1)}{\Delta H(0)} \times \frac{C_{dd}(0)}{C_{dd}(1)} \times \frac{C_{ff}(0)/C_{dd}(0)}{C_{ff}(1)/C_{dd}(1)}, \tag{2.7}$$

where the argument denotes the value of λ and for simplicity the time argument of the correlation functions has been omitted since it is always $t = 0$ in the QI approximation. Also, for convenience, we have again multiplied and divided both numerator and denominator by $C_{dd}(\lambda)$.

Four types of quantities have to be evaluated in order to compute the KIE from Eq. (2.7): the ratio of the partition functions $Q_r(1)/Q_r(0)$, ratio of the delta correlation functions $C_{dd}(1)/C_{dd}(0)$, and energy variance $\Delta H(\lambda)$ and the "velocity" factor $C_{ff}(\lambda)/C_{dd}(\lambda)$ for $\lambda = 0$ and 1. Since quantum statistical calculations are impractical on a grid for systems with more than a few degrees of freedom, one usually employs a path integral Monte Carlo approach. The last two of the four quantities are in the form of thermodynamic averages, and so can be computed directly by Metropolis Monte-Carlo techniques.[8,11,12] The first two quantities cannot be evaluated directly since they are ratios of quantities for two different values of λ.

Here is where considering a continuous isotope change becomes useful: instead of computing the ratios directly, we use a variant of *thermodynamic integration*,[17] applied to the

parameter λ (i.e., to the masses of the isotopes). The two ratios can be expressed as exponentials of the integrals of their logarithmic derivatives,

$$\frac{Q_r(1)}{Q_r(0)} = \exp\left[\int_0^1 d\lambda \frac{d\log Q_r(\lambda)}{d\lambda}\right], \tag{2.8}$$

$$\frac{C_{dd}(1)}{C_{dd}(0)} = \exp\left[\int_0^1 d\lambda \frac{d\log C_{dd}(\lambda)}{d\lambda}\right]. \tag{2.9}$$

Since the logarithmic derivatives are given by

$$\frac{d\log Q_r(\lambda)}{d\lambda} = \frac{dQ_r(\lambda)/d\lambda}{Q_r(\lambda)}, \tag{2.10}$$

they are normalized quantities (thermodynamic averages) which can be directly computed by the Metropolis algorithm. To summarize, we compute the ratios of both reactant partition functions and the delta-delta correlation functions at $\lambda = 0$ and 1 by computing the values of the corresponding logarithmic derivatives for enough values of λ between 0 and 1, and then by integrating over λ and exponentiating. The relevant path integral representations and estimators for $\Delta H(\lambda)$ and $C_{ff}(\lambda)/C_{dd}(\lambda)$ have been originally derived in Ref. 11. In our case, we need a generalized form which is written out explicitly in Appendix A of Ref. 8. Estimators for $d\log Q_r/d\lambda$ and $d\log C_{dd}/d\lambda$ needed in the thermodynamic integration are derived in the following section.

III. EFFICIENT ESTIMATORS FOR THE LOGARITHMIC DERIVATIVES

A. Path integral approach

Quantum statistical effects can be rigorously treated using the imaginary time path integral (PI). Let N be the number of atoms, d the number of spatial dimensions (usually $d = 3$), and P the number of imaginary time slices in the discretized PI ($P = 1$ gives classical mechanics, $P \to \infty$ gives quantum mechanics). Then the PI representation of the reactant partition function Q_r is

$$Q_r \simeq V^{-1} C \int d\mathbf{r}^{(0)} \int d\mathbf{r}^{(1)} \cdots \int d\mathbf{r}^{(P-1)} \rho_r\left(\left\{\mathbf{r}^{(s)}\right\}\right), \tag{3.1}$$

$$\rho_r\left(\left\{\mathbf{r}^{(s)}\right\}\right) = \exp\left[-\beta\Phi\left(\left\{\mathbf{r}^{(s)}\right\}\right)\right], \tag{3.2}$$

$$C \equiv \left(\frac{P}{2\pi\hbar^2\beta}\right)^{NdP/2} \prod_{i=1}^{N} m_i^{dP/2}. \tag{3.3}$$

where $\mathbf{r}^{(s)} \equiv \left(\mathbf{r}_1^{(s)}, \mathbf{r}_2^{(s)}, \ldots, \mathbf{r}_N^{(s)}\right)$ is the set of Cartesian coordinates associated with the sth time slice, V is the volume,[37] and $\Phi\left(\left\{\mathbf{r}^{(s)}\right\}\right)$ is the effective potential given by

$$\Phi\left(\left\{\mathbf{r}^{(s)}\right\}\right) = \frac{P}{2\hbar^2\beta^2}\sum_{i=1}^{N} m_i \sum_{s=0}^{P-1}\left(\mathbf{r}_i^{(s)} - \mathbf{r}_i^{(s+1)}\right)^2 + \frac{1}{P}\sum_{s=0}^{P-1} V\left(\mathbf{r}^{(s)}\right). \tag{3.4}$$

The presence of ρ_r in the path integral (3.1) strongly suppresses paths away from the reactant region because the quantum Boltzmann factor gives them an exponentially vanishing weight.

Similarly, the PI representation of C_{dd} is

$$C_{dd}(0) \simeq V^{-1} C \int d\mathbf{r}^{(1)} \int d\mathbf{r}^{(2)} \cdots \int d\mathbf{r}^{(P)} \rho^{\ddagger}\left(\left\{\mathbf{r}^{(s)}\right\}\right), \tag{3.5}$$

$$\rho^{\ddagger}\left(\left\{\mathbf{r}^{(s)}\right\}\right) = \Delta\left[\xi_a\left(\mathbf{r}^{(0)}\right)\right] \Delta\left[\xi_b\left(\mathbf{r}^{(P/2)}\right)\right] \exp\left[-\beta\Phi\left(\left\{\mathbf{r}^{(s)}\right\}\right)\right], \tag{3.6}$$

where $\rho^{\ddagger}$ is the thermal density matrix constrained to the two dividing surfaces ("transition state" region).[38] The dividing surfaces are defined by $\xi_\gamma(\mathbf{r}) = 0$, i.e., $\xi_a(\mathbf{r})$ and $\xi_b(\mathbf{r})$ are generalized reaction coordinates, functions of coordinates $\mathbf{r}$ that take on positive (negative) values on the product (reactant) sides of the dividing surfaces. Since the density $\rho^{\ddagger}$ contains the two delta constraints, the PI (3.5) selects paths that are close to the saddle point of the potential. In numerical applications, it is useful to replace the generalized delta function in $\rho^{\ddagger}$ by its Gaussian approximation.[8,11]

B. Thermodynamic estimators

By straightforward differentiation of (3.1) with respect to λ one obtains the "thermodynamic" estimator (TE)[39] for the logarithmic derivative (2.10) needed in expression (2.8),

$$\frac{d\log Q_r(\lambda)}{d\lambda} \simeq \sum_{i=1}^{N} \frac{dm_i}{d\lambda}\left\langle \frac{dP}{2m_i} - \frac{P}{2\hbar^2\beta}\sum_{s=0}^{P-1}\left(\mathbf{r}_i^{(s)} - \mathbf{r}_i^{(s+1)}\right)^2\right\rangle. \tag{3.7}$$

The similar thermodynamic estimator for the logarithmic derivative of C_{dd} is obtained by differentiating the logarithm of expression (3.5) and is given in Ref. 8.

These thermodynamic estimators have the advantage that they do not require evaluation of the potential. Their disadvantage, however, is that their statistical error σ_{stat} of the computed average grows with the number of slices P. If M denotes the number of Monte Carlo samples, τ the correlation time, then

$$\sigma_{\text{stat}} \approx O\left(P^{1/2}\tau^{1/2}M^{-1/2}\right). \tag{3.8}$$

C. Efficient virial estimators

In this subsection we present one of the central results of this paper, namely the efficient "virial"estimators for the logarithmic derivatives of Q_r and C_{dd}.

By rescaling the Cartesian coordinates in the PI (3.1) as $\mathbf{x}_i = m_i^{1/2}\mathbf{r}_i$, we obtain a PI in mass-scaled coordinates,

$$Q_r \simeq \left(\frac{P}{2\pi\hbar^2\beta}\right)^{NdP/2} \int d\mathbf{x}^{(0)} \int d\mathbf{x}^{(1)} \cdots \int d\mathbf{x}^{(P-1)} \exp\left[-\beta\Phi\left(\left\{\mathbf{x}^{(s)}\right\}\right)\right], \tag{3.9}$$

$$\Phi\left(\left\{\mathbf{x}^{(s)}\right\}\right) = \frac{P}{2\hbar^2\beta^2}\sum_{i=1}^{N}\sum_{s=0}^{P-1}\left(\mathbf{x}_i^{(s)} - \mathbf{x}_i^{(s+1)}\right)^2 + \frac{1}{P}\sum_{s=0}^{P-1} V\left(\left\{m_i^{-1/2}\mathbf{x}_i^{(s)}\right\}\right). \tag{3.10}$$

The logarithmic derivative (2.10) of PI (3.9) yields the simple virial estimator (SVE),

$$\frac{d\log Q_r}{d\lambda} \simeq -\frac{\beta}{P}\left\langle \sum_{s=0}^{P-1} \frac{dV\left[\left\{m_i\left(\lambda\right)^{-1/2}\mathbf{x}_i^{(s)}\right\}\right]}{d\lambda}\right\rangle \tag{3.11}$$

$$\simeq -\frac{\beta}{P}\left\langle \sum_{s=0}^{P-1} \frac{dV\left[\left\{m_i\left(\lambda+\Delta\lambda\right)^{-1/2} m_i(\lambda)^{1/2}\mathbf{r}_i^{(s)}\right\}\right]}{d\Delta\lambda}\right\rangle \tag{3.12}$$

$$\simeq \frac{\beta}{2P}\sum_{i=1}^{N}\frac{dm_i/d\lambda}{m_i}\left\langle \sum_{s=0}^{P-1}\mathbf{r}_i^{(s)}\cdot\frac{\partial V\left(\mathbf{r}^{(s)}\right)}{\partial\mathbf{r}_i^{(s)}}\right\rangle. \tag{3.13}$$

We have explicitly written three forms of the SVE. The first form (3.11) is for numerical evaluation if sampling is done in mass-scaled coordinates. (By "numerical" we mean that the total derivative of V with respect to λ is computed by finite difference.) The second form (3.12) is for numerical evaluation if sampling is done in Cartesian coordinates. Finally the third form (3.13) is for analytical evaluation when sampling is done in Cartesian coordinates. The second and third forms are especially useful since the estimators can be computed simultaneously with the TE (3.7). For PI Monte Carlo simulations, where forces are not computed in the simulation, the second form is the most useful one. In PI molecular dynamics simulations, one can equally well use the last estimator since the forces are available.

The statistical errors of the virial estimators, especially in systems with unbound degrees of freedom, can be drastically reduced by removing the centroid motion before rescaling,

i.e., by defining the mass-scaled coordinates as

$$\mathbf{x}_i = m_i^{1/2}(\mathbf{r}_i - \mathbf{r}_i^{(C)}),$$
$$\mathbf{r}_i^{(C)} = \frac{1}{P}\sum_{s=0}^{P-1}\mathbf{r}_i^{(s)}.$$

Following a more detailed derivation presented in Appendix A, one obtains the generalized virial estimator (GVE),

$$\frac{d\log Q}{d\lambda} \simeq \frac{d}{2}\sum_{i=1}^{N}\frac{dm_i/d\lambda}{m_i} - \frac{\beta}{P}\left\langle \sum_{s=0}^{P-1}\frac{dV\left[\left\{\mathbf{r}_i^{(C)} + m_i\left(\lambda\right)^{-1/2}\mathbf{x}_i^{(s)}\right\}\right]}{d\lambda}\right\rangle, \tag{3.14}$$

$$\simeq \frac{d}{2}\sum_{i=1}^{N}\frac{dm_i/d\lambda}{m_i} - \frac{\beta}{P}\left\langle \sum_{s=0}^{P-1}\frac{dV\left[\left\{\mathbf{r}_i^{(C)} + m_i\left(\lambda+\Delta\lambda\right)^{-1/2}m_i(\lambda)^{1/2}\left(\mathbf{r}_i^{(s)} - \mathbf{r}_i^{(C)}\right)\right\}\right]}{d\Delta\lambda}\right\rangle \tag{3.15}$$

$$\simeq \sum_{i=1}^{N}\frac{dm_i/d\lambda}{m_i}\left[\frac{d}{2} + \frac{\beta}{2P}\left\langle \sum_{s=0}^{P-1}\left(\mathbf{r}_i^{(s)} - \mathbf{r}_i^{(C)}\right)\cdot\frac{\partial V\left(\mathbf{r}^{(s)}\right)}{\partial \mathbf{r}_i^{(s)}}\right\rangle\right]. \tag{3.16}$$

The primary advantage of these estimators is that their statistical error is independent of the number P of imaginary time slices,

$$\sigma_{\text{stat}} \approx O\left(P^0\tau^{1/2}M^{-1/2}\right). \tag{3.17}$$

Again we have three forms (3.14), (3.15), and (3.16) for various applications: An analytical approach (3.16) would be useful in PI molecular dynamics simulations where gradient information is available. In PI Monte Carlo simulations where gradient is not needed for the random walk, computation of the gradient of the potential just for the estimators is costly, especially for polyatomic systems. Since we are only interested in the total derivative of the potential with respect to the change of a single parameter (λ), a direct computation of the total derivative by finite difference (3.15) is much more efficient. This approach requires at most $2P$ potential evaluations for each sampling and also we expect the (already very small) numerical errors of this approach to cancel under the Monte Carlo averaging. In practice the computational cost of evaluating virial estimators can be made negligible (relative to the cost of the random walk) by sampling only after several MC moves and by saving information about the previous sample and computing the new sample by evaluating the potential only for slices that moved in the present Monte Carlo step.

The virial estimator for the logarithmic derivative of C_{dd} is the same as for Q_r except that an additional term due to the constraint to the dividing surface is needed. This additional term is the same as for the thermodynamic estimator and can be found in Ref. 8. If the reader is confused by the notation of this section where general number of atoms and dimensions are used, a simplified version of the present approach with preliminary results was presented in Ref. 18, where the two types of estimators are derived and compared for a system with one degree of freedom (the Eckart barrier).

IV. KINETIC ISOTOPE EFFECTS IN *CIS*-PENTADIENE

In this section, the methodology described above is tested on the [1,5]-sigmatropic hydrogen shift in *cis*-1,3-pentadiene. This system has been studied by several authors both experimentally (by Roth and König[19]) and theoretically.[20–23] On one hand, this 13-atom system is much smaller than an enzyme; on the other hand, the pentadiene reaction exhibits certain properties believed to play a role in real enzymatic reactions.

The sigmatropic rearrangement involves an internal rotation to transform the reactant *s-trans* conformation to the transition state with C_s symmetry. The results presented here are for the same isotopologs that were measured experimentally: the KIE is the ratio of the rate constant k_{H} for D_2C-$(CH)_3$-CH_3 to that k_{D} for H_2C-$(CH)_3$-CD_3.

A. The potential energy surface for the reaction

There are several strategies for modeling a Born-Oppenheimer potential energy surface for a chemical reaction. The most accurate, but computationally most expensive, are *ab initio* quantum chemistry methods. Less accurate, but extremely efficient computationally, are the empirical *molecular mechanics force fields*. Since our aim is to develop methods applicable to systems as large as enzymes, it is necessary to use the second option.

Force fields, however, are designed to describe accurately only the stable chemical species near equilibrium, and are not suitable for modeling reactive systems, which requires accurate knowledge of the potential energy surface near the saddle point. On the other hand, if the properties of the reactive potential are known at least in the vicinity of the saddle point, there are various approaches for constructing a reactive potential from the molecular mechanics

force fields. Perhaps the most popular is the empirical valence bond (EVB) potential.[24,25]

The simplest EVB potential can be described as follows. One assumes that the reactants and products can be accurately described by some molecular mechanics potentials V_{11} and V_{22}, respectively. These are the diabatic potentials and the diagonal terms in the symmetric 2×2 EVB matrix, and the off-diagonal term V_{12} is the coupling between the two diabatic states. The EVB potential is the lower eigenvalue of the EVB matrix,

$$V_{\mathrm{EVB}} = \frac{1}{2}\left(V_{11} + V_{22}\right) - \sqrt{|V_{12}|^2 + \left(\frac{V_{11} - V_{22}}{2}\right)^2}. \tag{4.1}$$

In the simplest case, V_{12} is chosen to be a constant, for example to reproduce the barrier height of the EVB potential at the saddle point of an *ab initio* barrier height.

Constant coupling results in too strong mixing of V_{11} and V_{22} all the way to the reactant and product regions where the input potentials V_{11} and V_{22} are assumed to be accurate. An improvement is achieved by making the coupling decay, e.g., as a Gaussian function of the distance from the saddle point of the potential,

$$\begin{aligned} V_{12}(\mathbf{r}) &= A \exp(-\alpha |\Delta\mathbf{r}|^2/2), \\ \Delta\mathbf{r} &= \mathbf{r} - \mathbf{r}^{\ddagger} \end{aligned} \tag{4.2}$$

where $\mathbf{r}^{\ddagger}$ is the location of the saddle point of the potential (i.e., the transition state). This would probably be sufficient for getting an order of magnitude estimate for the rate constant, which depends most strongly on the barrier height. Since we are interested in the KIE, this information is not sufficient since, e.g., in the classical limit, the effect of the barrier height on the rate constant cancels out in the KIE ratio and the first nonvanishing effect on the KIE is due to the Hessian of the potential at the saddle point. A constant or a one-parameter Gaussian coupling cannot reproduce this Hessian.

Chang and Miller proposed a refinement of the EVB method which solved this problem.[25] In their approach, the EVB coupling has the form

$$V_{12}(\mathbf{r})^2 = A \exp(\mathbf{B} \cdot \Delta\mathbf{r} - \Delta\mathbf{r} \cdot \mathbf{C} \cdot \Delta\mathbf{r}),$$

and the constants A, $\mathbf{B}$, $\mathbf{C}$ are chosen in order to match the Hessian of the potential obtained by an *ab initio* calculation.

In our calculation, we used a further refinement by Schlegel and Sonnenberg,[26] in which the EVB coupling is given by

$$V_{12}(\mathbf{r}) = A[1 + \mathbf{B} \cdot \Delta\mathbf{r} + \Delta\mathbf{r} \cdot (\mathbf{C} + \alpha\mathbf{I}) \cdot \Delta\mathbf{r}] \exp(-\alpha|\Delta\mathbf{r}|^2/2). \tag{4.3}$$

The constants A, $\mathbf{B}$, $\mathbf{C}$ are again chosen to match the Hessian of the *ab initio* potential at the transition state. The advantage of this form is that it avoids divergences which occur sometimes in the Chang-Miller approach due to the fact that the quadratic form $\Delta\mathbf{r}\cdot\mathbf{C}\cdot\Delta\mathbf{r}$ is not guaranteed to be positive definite.

In our calculation, the *ab initio* Hessian of the potential for the pentadiene at the transition state was obtained at MP2/6-31g(d) level, using the Gaussian 03 software package.[27] For the nonreactive potentials V_{11} and V_{22} we chose the general AMBER force field (GAFF)[28] from the AMBER software package.[29] The GAFF potential has the form

$$\begin{aligned} V(\mathbf{r}) &= \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ &+ \sum_{\substack{\text{nonbonded} \\ \text{pairs } i<j}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_1 q_2}{\varepsilon r_{ij}} \right), \end{aligned} \tag{4.4}$$

in which the first term is due to bond stretches, the second due to bond bends, the third due to torsions, and the fourth due to Coulomb and Van der Waals interactions of nonbonded pairs of atoms. (See Ref. 28 for details.)

Two small modifications of the force field were necessary in order to prevent unreasonably large values of V_{11} near the saddle point (and the necessity of using a huge coupling V_{12}): First, the harmonic bond stretch potential for the bond being broken in the reaction was replaced by a more realistic Morse potential (the Morse potential can be fitted to the AMBER equilibrium bond length and force constant; the only additional parameter needed is the bond dissociation energy D). Second, the non-bonded interaction for the bond being formed in the reaction was omitted because the bond length in the transition state is much shorter than the equilibrium value for the Lennard-Jones potential and gives a diverging contribution to the potential.

As for the EVB potential (4.3), there is a free parameter α controlling the decay of the coupling. This α can be neither too large, because then the PES would not be smooth, nor too small, because then the reactant and product energies would be affected significantly. Ideally, there is an intermediate range of α for which the results are fairly independent of the precise value of α. Indeed this is the case for our pentadiene calculation, where this range is 0.3 a.u. $\lesssim \alpha \lesssim$ 1.5 a.u. We chose a value $\alpha = 0.9$ a.u.

The Monte Carlo simulation was performed in Cartesian coordinates. The question arises how to compute the distance in the configuration space of the current translated, rotated, and

deformed configuration from the configuration at the transition state. For a rigid body, this problem is fairly easy and is solved by translation and rotation. In the case of deformations, one has to invoke the so-called Eckart axis conditions[30]

$$\sum_{i=1}^{N} m_i \mathbf{r}_i^{\ddagger} \times (R\mathbf{r}_i) = 0. \tag{4.5}$$

where $\mathbf{r}_i$ are the current coordinates of the atoms, $\mathbf{r}_i^{\ddagger}$ are the coordinates at the transition state, and $R \in SO(3)$ is an appropriate rotation matrix that guarantees the Eckart axis conditions. In the above expression we for simplicity assumed that the center of mass was already subtracted from both current and the transition-state coordinates. It turns out that the Eckart axis conditions are equivalent to minimizing the distance between the two configurations in mass-scaled coordinates:[31]

$$\min_{R \in SO(3)} \sum_{i=1}^{N} m_i |\mathbf{r}_i^{\ddagger} - (R\mathbf{r}_i)|^2. \tag{4.6}$$

In the actual calculation, one has to find the appropriate rotation matrix R after each Monte Carlo step in order to minimize the distance or, equivalently, satisfy the Eckart axis conditions. We used the algorithm of Dymarsky and Kudin.[32]

B. Details of the calculation

We computed the KIE for the intramolecular hydrogen transfer in *cis*-pentadiene at four temperatures in the range 463-500K, as in the experiment by Roth and König.[19] As opposed to various types of mixed quantum-classical treatments in the literature, in our calculation all atoms were treated quantum-mechanically, using P imaginary time slices in the path integral. In order to check the convergence of the results, P was changed from 4 to 50, and it was found that $P = 24$ suffices. For each value of P, a Monte Carlo simulation was run in which four quantities were computed: the ratio of the partition functions $Q_r(1)/Q_r(0)$, the ratio of the delta-delta correlation functions $C_{dd}(1)/C_{dd}(0)$, the energy variance $\Delta H(\lambda)$ and the "velocity" factor $C_{ff}(\lambda)/C_{dd}(\lambda)$ for $\lambda = 0$ and 1. Ratios $Q_r(1)/Q_r(0)$ and $C_{dd}(1)/C_{dd}(0)$ were computed by thermodynamic integration (2.8) and (2.9) over five equally spaced values of λ between 0 and 1. The logarithmic derivatives, in turn, were computed using both thermodynamic and virial estimators from Sec. III.

Similarly as in Refs. 11 and 8, the random walk was implemented using the Metropolis Monte Carlo algorithm for path integrals and the staging algorithm.[33] The total number of Monte Carlo samples for the logarithmic derivatives was 2×10^6 for Q_r and 16×10^6 for C_{dd}, regardless of P. Since only thermodynamic estimators were used for $\Delta H(\lambda)$ and $C_{ff}(\lambda)/C_{dd}(\lambda)$, the number of samples was increased with P as $P \times 1.6 \times 10^6$, in order to keep the statistical error of the QI expression for the KIE within 3%. In order to efficiently sample the configuration space, three types of Monte Carlo moves were used: a single-slice move, a multi-slice move, and a whole-chain move, with respective frequencies: 0.50, 0.45, and 0.05.[11]

C. Numerical results

The result of the thermodynamic integration (2.8) and (2.9) for partition functions and delta-delta correlation functions are shown in Figs. 1 and 2. Both figures compare results calculated with the thermodynamic and virial estimators from Eqs. (3.7) and (3.15), respectively. Figure 1 a) shows the convergence of $Q_r(1)/Q_r(0)$ as a function of the number P of imaginary time slices. Part b) shows the P-dependence of the relative statistical error. It is clear that the virial estimator (3.15) is superior to the thermodynamic estimator (3.7) from Ref. 8: While the error is constant for the virial estimator, it grows with P for the thermodynamic estimator.

Similar results can be seen in Fig. 2 that shows the P dependence of the ratio of delta-delta correlation functions and its statistical error. In order to justify that the thermodynamic integration is good enough with only five values of λ, Fig. 3 shows the λ dependence of the logarithmic derivative of Q_r and C_{dd}. Both graphs are fairly smooth, in fact almost linear functions, and therefore the error introduced by the discretization of the integral is small.

Finally, the KIE itself is displayed in Fig. 4 and Table I. The temperature dependence of the QI result is compared with the experimental results by Roth and König,[19] conventional transition state theory (TST) calculations by Dormans and Buck,[20] and the canonical variational TST with multidimensional SC tunneling (CVT/SCT) by Liu et al.[22] Calculation by Dormans and Buck used Hessian information about the potential at the reactant minimum and at the saddle point obtained by ab initio calculations at the 3-21G level. Liu et al. used a semiempirical AM1 potential.

As expected, the conventional TST underestimates the KIE by not treating quantum effects properly. The variational TST with SC tunneling seems to match experimental results quite well. In fact the agreement is almost surprising since only a semiempirical potential is used.

It should be noted that the QI results are somewhat dependent on the precise value of α used in the EVB potential (4.3). Depending on α, the KIE can vary by about 10%. This is a fairly low error, considering that the potential was based on force fields at the reactant and product wells, and only fitted to the *ab initio* Hessian at the transition state. What is also satisfactory is that the temperature dependence of the KIE is very close to the experimental dependence. It appears that the QI takes correctly into account the quantum effects on the KIE. The small remaining error is mostly due to the accuracy of the PES and partially due to the recrossing effects.

V. DISCUSSION

This paper has extended the methodology of computing KIE's using the quantum instanton approximation from Ref. 8. New virial estimators for the logarithmic derivatives needed in the thermodynamic integration have been derived and shown to be much more efficient than the original thermodynamic estimators used in Ref. 8. In future applications, the generalized virial estimators will be the estimators of choice because their statistical errors are independent of number P of imaginary time slices. The improved approach was applied to the [1,5]-sigmatropic hydrogen shift in *cis*-1,3-pentadiene. The results of the calculation showed the importance of quantum effects in the reaction.

There are several possible extensions that will be useful in applications of this methodology to large biochemical reactions. First, the calculation can be sped up by using a different number of imaginary time slices for different degrees of freedom: fewer slices would be used for heavier, more classical atoms or atoms far from the active site of the enzyme.[14,34] Second, a more general approach for systems with no symmetry can use dividing surfaces orthogonal to the numerically obtained minimum energy path. In fact, this somewhat simplifies the estimators for constrained quantities and should also minimize recrossing effects.[35] Third, because of the expense of *ab initio* or even semiempirical potentials, for large systems the EVB potential should be used. However, the EVB potential should be improved by incor-

porating information about the Hessian of the reactive potential at several points along the reaction path obtained by *ab initio* calculations (rather than just at the saddle point).[26] Finally, more efficient virial estimators should be used also for the calculation of $\Delta H(\lambda)$ and $C_{ff}(\lambda)/C_{dd}(\lambda)$.[36]

The main advantages of the PI approach using QI approximation are that it does not use semiclassical approximation for the Boltzmann operator and that it does not arbitrarily separate degrees of freedom into classical and quantum, which is done in mixed quantum-classical approaches. To conclude, the QI approximation combined with thermodynamic integration, path integral Metropolis Monte Carlo implementation, and the efficient virial estimators derived in this paper appears to be a promising approach to compute kinetic isotope effects in polyatomic systems.

Acknowledgments

This work was supported by the Office of Naval Research Grant No. N00014-05-1-0457 and by the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We would like to thank H. B. Schlegel and J. L. Sonnenberg for discussions and help with the Gaussian 03 software package.

APPENDIX A: DERIVATION OF THE GENERALIZED VIRIAL ESTIMATOR

The easiest way to derive this estimator is by using the mass-scaled coordinates for the path integrals. The dependence on mass is therefore shifted from the kinetic part to the potential part of the action. Starting from the PI (3.1), we change coordinates to $\mathbf{y}^{(s)} \equiv \mathbf{r}^{(s)} - \mathbf{r}^{(0)}$ for $s = 1, \ldots, P-1$, (i.e., we subtract coordinates of one of the slices from the rest) and obtain

$$Q_r \simeq C \int d\mathbf{r}^{(0)} \int d\mathbf{y}^{(1)} \cdots \int d\mathbf{y}^{(P-1)} e^{-\beta\Phi},$$

$$\Phi = \frac{P}{2\hbar^2\beta^2} \sum_{i=1}^{N} m_i \left[\left(\mathbf{y}_i^{(1)}\right)^2 + \sum_{s=1}^{P-2} \left(\mathbf{y}_i^{(s)} - \mathbf{y}_i^{(s+1)}\right)^2 + \left(\mathbf{y}_i^{(P-1)}\right)^2 \right] + \frac{1}{P} \left[V\left(\mathbf{r}^{(0)}\right) + \sum_{s=1}^{P-1} V\left(\mathbf{r}^{(0)} + \mathbf{y}^{(s)}\right) \right].$$

Now we define mass-scaled coordinates as $\mathbf{x}_i^{(s)} \equiv m_i^{1/2}\mathbf{y}_i^{(s)}$ and obtain

$$Q_r \simeq \left(\frac{P}{2\pi\hbar^2\beta}\right)^{NdP/2} \prod_{i=1}^{N} m_i^{d/2} \int d\mathbf{r}^{(0)} \int d\mathbf{x}^{(1)} \cdots \int d\mathbf{x}^{(P-1)} e^{-\beta\Phi}, \tag{A1}$$

$$\begin{aligned}\Phi &= \frac{P}{2\hbar^2\beta^2} \sum_{i=1}^{N} \left[\left(\mathbf{x}_i^{(1)}\right)^2 + \sum_{s=1}^{P-2} \left(\mathbf{x}_i^{(s)} - \mathbf{x}_i^{(s+1)}\right)^2 + \left(\mathbf{x}_i^{(P-1)}\right)^2 \right] \\ &\quad + \frac{1}{P} \left[V\left(\mathbf{r}^{(0)}\right) + \sum_{s=1}^{P-1} V\left(\{\mathbf{r}_i^{(0)} + m_i^{-1/2}\mathbf{y}_i^{(s)}\}\right) \right].\end{aligned} \tag{A2}$$

Taking the logarithmic derivative of Eq. (A1), we find a set of equations equivalent to Eqs. (3.14)-(3.16), except that $\mathbf{r}_i^{(C)}$ should be replaced by $\mathbf{r}_i^{(0)}$. E.g., instead of Eq. (3.16) we obtain

$$\frac{d\log Q_r}{d\lambda} \simeq \sum_{i=1}^{N} \frac{dm_i/d\lambda}{m_i} \left[\frac{d}{2} + \frac{\beta}{2P} \left\langle \sum_{s=0}^{P-1} \left(\mathbf{r}_i^{(s)} - \mathbf{r}_i^{(0)}\right) \cdot \frac{\partial V\left(\mathbf{r}^{(s)}\right)}{\partial \mathbf{r}_i^{(s)}} \right\rangle \right]. \tag{A3}$$

Since we have chosen the slice $s = 0$ arbitrarily, we can do the same with all slices $s = 0, \ldots, P-1$, obtain analogs of estimator (A3) and take the average of these estimators. We thus obtain the GVE of Eq. (3.16), and working backward also the variants (3.14) and (3.15).

---

* Electronic address: `vanicek@ias.edu`

[1] L. C. S. Melander, *Isotope effects on reaction rates* (Ronald Press Co., New York, 1960).

[2] A. Kohen, R. Cannio, S. Bartolucci, and J. P. Klinman, Nature **399**, 496 (1999).

[3] J. Basran, M. J. Sutcliffe, and N. S. Scrutton, Biochemistry **38**, 3218 (1999).

[4] T. D. Poulsen, M. Garcia-Viloca, J. Gao, and D. G. Truhlar, J. Phys. Chem. B **107**, 9567 (2003).

[5] Z. Smedarchina, W. Siebrand, A. Fernandez-Ramos, and Q. Cui, J. Am. Chem. Soc. **125**, 243 (2003).

[6] S.-Y. Kim and S. Hammes-Schiffer, J. Chem. Phys. **119**, 4389 (2003).

[7] M. H. M. Olsson, P. E. M. Siegbahn, and A. Warshel, J. Am. Chem. Soc. **126**, 2820 (2003).

[8] J. Vaníček, W. H. Miller, J. F. Castillo, and F. J. Aoiz, J. Chem. Phys. **123**, 054108 (2005).

[9] W. H. Miller, Y. Zhao, M. Ceotto, and S. Yang, J. Chem. Phys. **119**, 1329 (2003).

[10] W. H. Miller, J. Chem. Phys. **62**, 1899 (1975).

[11] T. Yamamoto and W. H. Miller, J. Chem. Phys. **120**, 3086 (2004).

[12] Y. Zhao, T. Yamamoto, and W. H. Miller, J. Chem. Phys. **120**, 3100 (2004).

[13] W. H. Miller, M. Ceotto, and S. Yang, J. Chem. Phys. **122**, 044109 (2005).

[14] Y. Li and W. H. Miller, Mol. Phys. **103**, 203 (2004).

[15] M. Ceotto and W. H. Miller, not published.

[16] W. H. Miller, S. D. Schwartz, and J. W. Tromp, J. Chem. Phys. **79**, 4889 (1983).

[17] D. Chandler, *Introduction to modern statistical mechanics* (Oxford University Press, New York, 1987).

[18] J. Vaníček and W. H. Miller, In *Proceedings of the 8th International Conference: Path Integrals from Quantum Information to Cosmology* ed. C. Burdik, O. Navratil, and S. Posta (JINR, Dubna, 2005).

[19] W. R. Roth and J. König, J. Liebigs. Ann. Chem. **24**, 699 (1966).

[20] G. J. M. Dormans and H. Buck, J. Am. Chem. Soc. **108**, 3253 (1986).

[21] L. Chantranupong and T. A. Wildman, J. Am. Chem. Soc. **112**, 4151 (1990).

[22] Y.-P. Liu, G. C. Lynch, T. N. Truong, D. Lu, D. G. Truhlar, and B. C. Garrett, J. Am. Chem. Soc. **115**, 2408 (1993).

[23] L. J. Schaad, L. Bytautas, and K. N. Houk, Can. J. Chem. **77**, 875 (1999).

[24] A. Warshel and R. M. Weiss, J. Am. Chem. Soc. **102**, 6218 (1980).

[25] Y.-T. Chang and W. H. Miller, J. Phys. C **94**, 5884 (1990).

[26] H. B. Schlegel and J. L. Sonnenberg, J. Chem. Theory Comput. **2**, 905 (2006).

[27] M. J. F. et al.", *Gaussian 03* (Gaussian, Inc., Wallingford, CT, 2004).

[28] J. Wang, R. M. Wolf, J. W. Caldwell, P. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[29] D. A. Case, I. T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, J. K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods, J. Comput. Chem. **26**, 1668 (2005).

[30] C. Eckart, Phys. Rev. **47**, 552 (1935).

[31] K. N. Kudin and A. Y. Dymarsky, J. Chem. Phys. **122**, 224105 (2005).

[32] A. Y. Dymarsky and K. N. Kudin, J. Chem. Phys. **122**, 124103 (2005).

[33] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).

[34] T. Yamamoto and W. H. Miller, J. Chem. Phys. **122**, 044106 (2005).

[35] Y. Li and W. H. Miller (2006), in preparation.

[36] S. Yang, T. Yamamoto, and W. H. Miller, J. Chem. Phys. **124**, 084102 (2006).

[37] This expression is for a bimolecular reaction. In the limit $V \to \infty$, V^{-2} cancels the V^2 divergence from the integration over the center-of-mass coordinate and the relative coordinate of the two molecules. For a unimolecular reaction, the factor is V^{-1} and is canceled by integration over the center-of-mass coordinate.

[38] Factor V^{-1} is canceled by the integral over the center of mass coordinate. Expression (3.5) is valid for both unimolecular and bimolecular reactions.

[39] The name comes from an analogy with an estimator for the kinetic energy derived in a similar manner.

TABLES

TABLE I: Experimental and theoretical kinetic isotope effect in *cis*-1,3-pentadiene for several temperatures.

T(K)	exper.[19]	TST (3-21G)[20]	CVT/SCT (AM1)[22]	QI (EVB + MP2/6-31g(d))
463.25	5.3	2.51	5.2	5.4
470	5.2	2.48	4.9	5.3
478.45	5.0	2.45	4.7	5.1
500	4.7	2.36	4.1	4.3

FIGURE CAPTIONS

FIG. 1: a) Dependence of the partition function ratio on the number P of imaginary time slices. b) P dependence of the statistical error.

FIG. 2: a) Dependence of the delta-delta correlation function ratio on the number P of imaginary time slices. b) P dependence of the statistical error.

FIG. 3: Input for thermodynamic integration: a) λ dependence of the logarithmic derivative of the partition function. b) λ dependence of the logarithmic derivative of the delta-delta correlation function.

FIG. 4: Temperature dependence of the experimental and computed kinetic isotope effect.

FIGURES

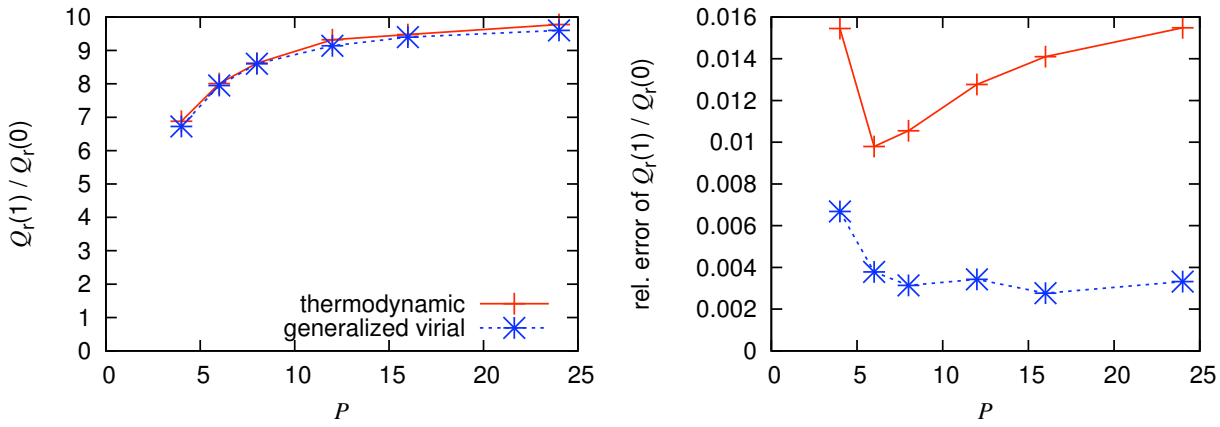

FIG. 1: J. Vanicek et al.

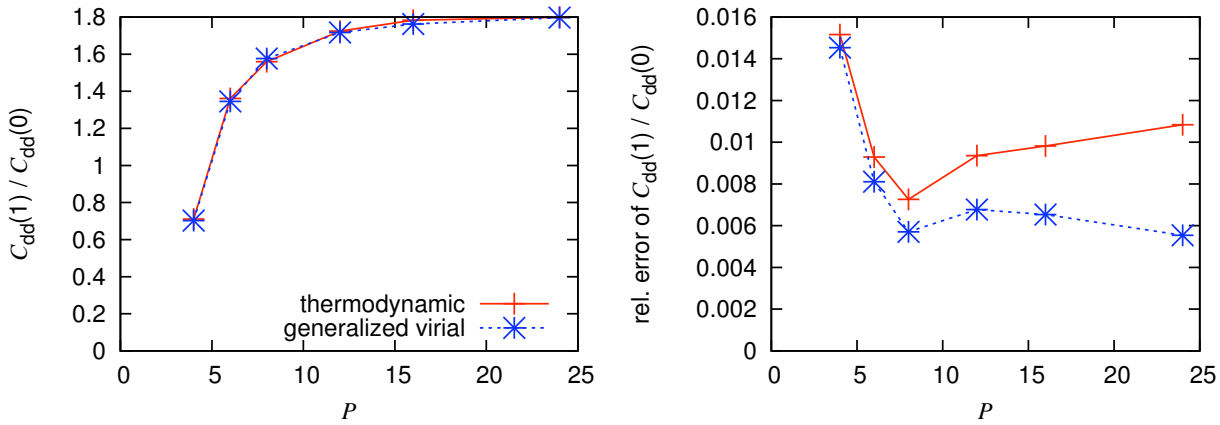

FIG. 2: J. Vanicek et al.

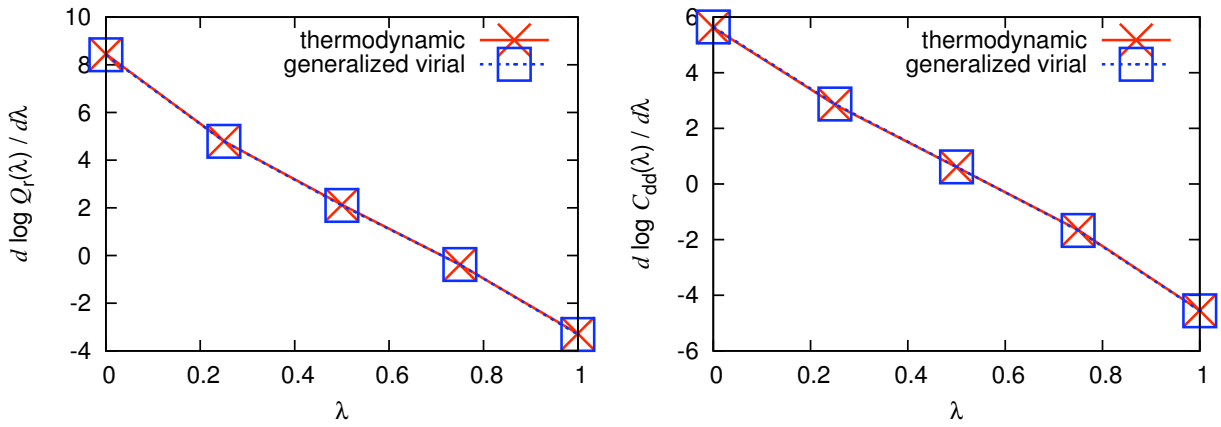

FIG. 3: J. Vanicek et al.

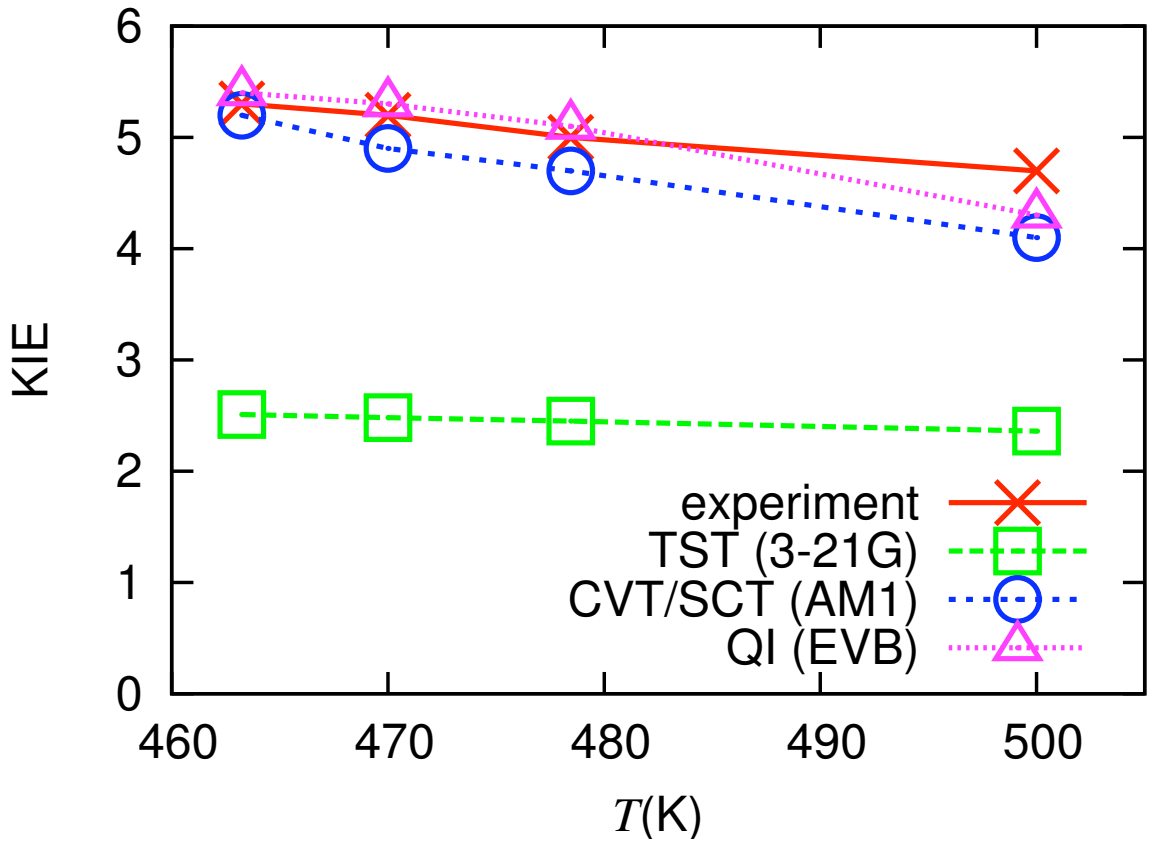

FIG. 4: J. Vanicek et al.